

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

4 日本国特許庁

PATENT OFFICE  
JAPANESE GOVERNMENT

02.06.00

JP00/3623

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日  
Date of Application:

1999年 6月 4日

27 JUL 2000

出願番号  
Application Number:

平成11年特許願第158497号

出願人  
Applicant(s):

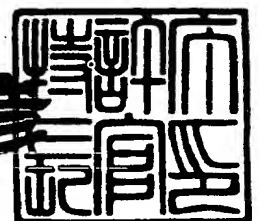
セイコーエプソン株式会社

PRIORITY  
DOCUMENT  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

2000年 6月29日

特許庁長官  
Commissioner,  
Patent Office

近藤隆彦



出証番号 出証特2000-3052005

【書類名】 特許願

【整理番号】 J0074153

【提出日】 平成11年 6月 4日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 17/30  
G06F 7/36

【発明の名称】 情報分類方法及び情報分類装置並びに情報分類処理プログラムを記録した記録媒体

【請求項の数】 21

【発明者】

【住所又は居所】 長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社内

【氏名】 長石 道博

【発明者】

【住所又は居所】 長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社内

【氏名】 三輪 真司

【特許出願人】

【識別番号】 000002369

【氏名又は名称】 セイコーエプソン株式会社

【代表者】 安川 英昭

【代理人】

【識別番号】 100093388

【弁理士】

【氏名又は名称】 鈴木 喜三郎

【連絡先】 0 2 6 6 - 5 2 - 3 1 3 9

【選任した代理人】

【識別番号】 100095728

【弁理士】

【氏名又は名称】 上柳 雅誉

【選任した代理人】

【識別番号】 100107261

【弁理士】

【氏名又は名称】 須澤 修

【手数料の表示】

【予納台帳番号】 013044

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9711684

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 情報分類方法及び情報分類装置並びに情報分類処理プログラムを記録した記録媒体

【特許請求の範囲】

【請求項 1】 検索サービスで検索された複数の検索結果をクラスタリングモジュールが取得して、そのクラスタリングモジュールでは、その検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を出力することを特徴とする情報分類方法。

【請求項 2】 前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュールを備えたことを特徴とする請求項 1 記載の情報分類方法。

【請求項 3】 前記変換モジュールは、クラスタリング処理を複数の検索サービスに対応して行う場合、複数の検索サービスそれぞれに対応して設けられることを特徴とする請求項 2 記載の情報分類方法。

【請求項 4】 前記複数の検索サービスのいずれかを選択的に用いて検索処理を行い、選択された検索サービスで検索された検索結果に対してクラスタリング処理を行うことを特徴とする請求項 3 記載の情報分類方法。

【請求項 5】 前記複数の検索サービスの少なくとも 2 つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果を収集して、その収集された検索結果に対してクラスタリング処理を行うことを特徴とする請求項 3 記載の情報分類方法。

【請求項 6】 前記複数の検索サービスの少なくとも 2 つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果に対し個別にクラスタリング処理を行うことを特徴とする請求項 3 記載の情報分類方法。

【請求項 7】 前記検索結果に対しクラスタリング処理を行う際、クラスタリング対象とする情報は、それぞれの検索結果における文書のタイトル、URL アドレス、更新日時、ファイルサイズの少なくとも 1 つであることを特徴とする請求項 1 から 6 のいずれか 1 項に記載の情報分類方法。

【請求項 8】 検索サービスで検索された複数の検索結果を入力し、入力された複数の検索結果に対しクラスタリング処理するクラスタリングモジュールを有し、このクラスタリングモジュールは、前記検索サービスで検索された複数の検索結果を取得して、その検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を出力することを特徴とする情報分類装置。

【請求項 9】 前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュールを備えたことを特徴とする請求項 8 記載の情報分類装置。

【請求項 10】 前記変換モジュールは、クラスタリング処理を複数の検索サービスに対応して行う場合、複数の検索サービスそれぞれに対応して設けられることを特徴とする請求項 9 記載の情報分類装置。

【請求項 11】 前記複数の検索サービスのいずれかを選択的に用いて検索処理を行い、選択された検索サービスで検索された検索結果に対してクラスタリング処理を行うことを特徴とする請求項 10 記載の情報分類装置。

【請求項 12】 前記複数の検索サービスの少なくとも 2 つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果を収集してまとめる検索結果収集部を設け、そのまとめられた検索結果に対してクラスタリング処理を行うことを特徴とする請求項 10 記載の情報分類装置。

【請求項 13】 前記複数の検索サービスの少なくとも 2 つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果に対し個別にクラスタリング処理を行うことを特徴とする請求項 10 記載の情報分類装置。

【請求項 14】 前記検索結果に対しクラスタリング処理を行う際、クラスタリング対象とする情報は、それぞれの検索結果における文書のタイトル、URL アドレス、更新日時、ファイルサイズの少なくとも 1 つであることを特徴とする請求項 8 から 13 のいずれか 1 項に記載の情報分類装置。

【請求項 15】 検索サービスで検索されたユーザの検索要求に対する複数の検索結果をクラスタリングモジュールがクラスタリング処理してそのクラスタリング処理結果を出力する情報分類処理プログラムを記録した記録媒体であって、その情報処理プログラムは、検索サービスからの検索結果を取得する手順と、

取得した検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を出力する手順とを含むことを特徴とする情報分類処理プログラムを記録した記録媒体。

【請求項 16】 前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する手順を行ったのちにクラスタリング処理を行うことを特徴とする請求項 15 記載の情報分類処理プログラムを記録した記録媒体。

【請求項 17】 前記クラスタリング処理を複数の検索サービスに対応して行う場合、前記クラスタリングモジュールが処理可能な形式に変換する手順は複数の検索サービスそれぞれに対応して行うことを特徴とする請求項 16 記載の情報分類処理プログラムを記録した記録媒体。

【請求項 18】 複数の検索サービスのいずれかを選択的に用いて検索処理を行い、選択された検索サービスで検索された検索結果に対してクラスタリング処理を行うことを特徴とする請求項 17 記載の情報分類処理プログラムを記録した記録媒体。

【請求項 19】 前記複数の検索サービスの少なくとも 2 つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果を収集して、その収集された検索結果に対してクラスタリング処理を行うことを特徴とする請求項 17 記載の情報分類処理プログラムを記録した記録媒体。

【請求項 20】 前記複数の検索サービスの少なくとも 2 つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果に対し個別にクラスタリング処理を行うことを特徴とする請求項 17 記載の情報分類処理プログラムを記録した記録媒体。

【請求項 21】 前記検索結果に対しクラスタリング処理を行う際、クラスタリング対象とする情報は、それぞれの検索結果における文書のタイトル、URL アドレス、更新日時、ファイルサイズの少なくとも 1 つであることを特徴とする請求項 15 から 20 のいずれか 1 項に記載の情報分類処理プログラムを記録した記録媒体。



## 【発明の詳細な説明】

## 【0001】

## 【発明の属する技術分野】

本発明は汎用の検索サービスで検索された結果に対しクラスタリング処理を施すことで、ユーザに見易い形で提示するようにした情報分類方法及び情報分類装置並びに情報分類処理プログラムを記録した記録媒体に関する。

## 【0002】

## 【従来の技術】

ネットワーク上に存在する膨大な量の情報の中からユーザの所望とする情報を検索する場合、検索サービスの存在は重要である。たとえば、インターネットでwebページを検索する際、ユーザは、幾つかの検索サービスの中から任意の検索サービスを選び、自分の欲しい情報を得るための検索要求としてのキーワードを入力する。これによって、検索サービス側では、入力されたキーワードに基づいて情報検索を行って、その検索結果をユーザに提示する。

## 【0003】

しかし、検索サービスによって検索される情報は膨大な量となることも多く、その中からユーザの本当に欲しい情報を見つけるのは非常に大変である。近年、webページは増大の一途をたどっているため、検索された多数の情報を如何にユーザにわかりやすく提示するかが大きな課題となっている。

## 【0004】

最近では、検索された情報をユーザが見やすい形に加工して提示する手法も開発され実用化されつつある。たとえば、ユーザの入力したキーワードで検索された結果から得られるキーワードを用いて再検索することで、情報の絞り込みを行い、ユーザの所望とするwebページを見つけやすくする方法がある。つまり、検索によって得られる検索結果の集合を特徴づけるキーワードを抽出して、ユーザの本当に欲しい情報の集合に収束させる方法である。

## 【0005】

## 【発明が解決しようとする課題】

このように、膨大な情報の中から、似た情報の集合を見つけることをクラスタ

リングという。情報処理ではこのクラスタリングはよく知られた手法であり、膨大な文書を分類する場合などに一般に使われている。

【0006】

しかし、現在、一般のユーザに広く利用されている検索サービス（汎用検索サービスという）が検索した結果をクラスタリングすることは行われてはなく、前述したように、入力されたキーワードに基づいて情報を抽出して、抽出された情報を羅列的にユーザに提示するというのが一般的である。したがって、ユーザは、その羅列された多数の情報から自分の欲しい情報を探すという面倒な作業をせざるを得なかった。

【0007】

そこで本発明は、汎用の検索サービスで得られた検索結果をクラスタリングすることで、ユーザに対し見易い形での検索情報提示を可能とすることを目的としている。

【0008】

【課題を解決するための手段】

前述の目的を達成するために、本発明の情報分類方法は、検索サービスで検索された複数の検索結果をクラスタリングモジュールが取得して、そのクラスタリングモジュールでは、その検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を出力するようにしている。

【0009】

この情報分類方法において、前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュールを備えるようにしている。

【0010】

そして、前記変換モジュールは、クラスタリング処理を複数の検索サービスに対応して行う場合、複数の検索サービスそれぞれに対応して設けられる。

【0011】

このとき、複数の検索サービスのいずれかを選択的に用いて検索処理を行い、選択された検索サービスで検索された検索結果に対してクラスタリング処理を行

うことも可能であり、また、前記複数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果を収集して、その収集された検索結果に対してクラスタリング処理を行うことも可能であり、さらに、前記複数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果に対し個別にクラスタリング処理を行うことも可能である。

【0012】

また、前記検索結果に対しクラスタリング処理を行う際、クラスタリング対象とする情報は、それぞれの検索結果における文書のタイトル、URLアドレス、更新日時、ファイルサイズの少なくとも1つとしている。

【0013】

また、本発明の情報分類装置は、検索サービスで検索された複数の検索結果を入力し、入力された複数の検索結果に対しクラスタリング処理するクラスタリングモジュールを有し、このクラスタリングモジュールは、前記検索サービスで検索された複数の検索結果を取得して、その検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を出力するようにしている。

【0014】

この情報分類装置において、前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュールを備えるようにしている。

【0015】

そして、前記変換モジュールは、クラスタリング処理を複数の検索サービスに対応して行う場合、複数の検索サービスそれぞれに対応して設けられる。

【0016】

このとき、複数の検索サービスを選択的に用いて検索処理を行い、選択された検索サービスで検索された検索結果に対してクラスタリング処理を行うことも可能であり、また、前記複数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果を収集して、その収集された検索結果に対してクラスタリング処理を行うことも可能であり、さらに、前記複

数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果に対し個別にクラスタリング処理を行うことも可能である。

## 【0017】

また、前記検索結果に対しクラスタリング処理を行う際、クラスタリング対象とする情報は、それぞれの検索結果における文書のタイトル、URLアドレス、更新日時、ファイルサイズの少なくとも1つとしている。

## 【0018】

また、本発明の情報分類処理プログラムを記録した記録媒体は、検索サービスで検索されたユーザの検索要求に対する複数の検索結果をクラスタリングモジュールがクラスタリング処理してそのクラスタリング処理結果を提示する情報分類処理プログラムを記録した記録媒体であって、その情報処理プログラムは、検索サービスからの検索結果を取得する手順と、取得した検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を出力する手順とを含むものである。

## 【0019】

この情報分類処理プログラムを記録した記録媒体において、前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する手順を行ったのちにクラスタリング処理を行うようにしている。

## 【0020】

そして、前記クラスタリング処理を複数の検索サービスに対応して行う場合、前記クラスタリングモジュールが処理可能な形式に変換する手順は複数の検索サービスそれぞれに対応して行うようにしている。

## 【0021】

このとき、複数の検索サービスのいずれかを選択可能とし、選択された検索サービスで検索された検索結果に対してクラスタリング処理を行うことも可能であり、また、前記複数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果を収集して、その収集された検索結果に対してクラスタリング処理を行うことも可能であり、さらに、前記複数の検

索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果に対し個別にクラスタリング処理を行うことも可能である。

【0022】

また、前記検索結果に対しクラスタリング処理を行う際、クラスタリング対象とする情報は、それぞれの検索結果における文書のタイトル、URLアドレス、更新日時、ファイルサイズの少なくとも1つとしている。

【0023】

このように本発明は、検索サービスで検索された多数の検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を提示するようにしているので、従来の検索結果をそのまま羅列的に出力するものと異なり、ユーザが見易いように整理されたものとなる。

【0024】

なお、この場合、検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュールを備える。そして、その変換モジュールを複数の検索サービスそれぞれに対応して用意することによって、複数の検索サービスで検索された検索結果に対してクラスタリング処理が可能となる。

【0025】

たとえば、複数の検索サービスの中から選択された検索サービスで検索された検索結果に対してクラスタリング処理が可能となる。これによって、ユーザは、情報を検索しようとする際、検索しようとする情報の種類によって最適な検索サービスを選ぶことができる。さらに、1つの検索サービスが混雑している場合、他の検索サービスに切り換えて検索を行うということも可能となる。

【0026】

また、複数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、複数の検索サービスで得られたそれぞれの検索結果を収集して、その収集された検索結果に対してクラスタリング処理を行うことができる。これによって、広範囲にわたって存在する情報を効率よく検索することが可能となり、その広範囲にわたる情報の検索結果を1つにまとめ、そのまとめられた結果

をクラスタリングすることができるので、多種多様な情報のクラスタリング結果を提示できる。

【0027】

さらに、複数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索サービスで検索された情報を個々にクラスタリングしてそれぞれのクラスタリングを提示することも可能で、これによれば、広範囲にわたって存在する情報を効率よく検索することが可能となり、また、それぞれのクラスタリング結果を比較することが可能となり、それぞれの検索サービスの特色などを知ることができる。

【0028】

また、クラスタリング対象とする情報は、それぞれの検索結果における文書のタイトル、URLアドレス、更新日時、ファイルサイズの少なくとも1つとして、いる。これによって、クラスタリングを文書全体から行う場合に比べて、クラスタリング処理を簡単なものとすることができ、それぞれに特色を有したクラスタリング結果が得られる。

【0029】

#### 【発明の実施の形態】

以下、本発明の実施の形態について説明する。なお、この実施の形態で説明する内容は、本発明の情報分類方法および情報分類装置についての説明であるとともに、本発明の情報分類処理プログラムを記録した記録媒体における情報分類処理プログラムの具体的な処理内容をも含むものである。

【0030】

図1は本発明を実現するための装置構成を示すもので、大きく分けると、検索サービス1、変換モジュール2、クラスタリングモジュール3とから構成され、変換モジュール2とクラスタリングモジュール3が本発明の情報分類装置に相当する。

【0031】

検索サービス1はインターネットなどで一般に広く使用されている汎用の検索サービスであり、ユーザからの検索要求としてのキーワードが入力されることに

より、たとえばwebページなどから、入力されたキーワードに基づいた情報検索を行う。この検索サービス1で検索された検索結果はファイル形式で出力され、クラスタリングモジュール3に渡されるが、検索サービス1は複数存在し、それぞれの検索サービスによって出力されるデータ形式が異なるので、どのような検索サービスからのファイルであってもその内容を読めるような形式にファイルを変換するために変換モジュール2が設けられる。

#### 【0032】

クラスタリングモジュール3は、検索サービス1により出力された検索結果ファイル内容（変換モジュール1による変換後のファイル内容）からクラスタリングするために必要な情報（クラスタリング対象情報という）を抽出するクラスタリング対象情報抽出部31、そのクラスタリング対象情報として抽出された情報を形態素解析する形態素解析部32、その形態素解析結果に基づいてクラスタリングを行うクラスタリング処理部33などを有している。

#### 【0033】

クラスタリング対象情報抽出部31は、変換モジュール部2で変換された検索エンジン1の検索結果から、クラスタリング対象情報を抽出するものであり、このクラスタリング対象情報としては幾つか考えられる（後述する）。この実施の形態では、検索結果として抽出された多数の文書のそれぞれのタイトル（見出し）をクラスタリング対象情報として抽出するものとする。たとえば、検索結果として図2に示すような複数の文書D1, D2, ..., D7が得られたとする。これら文書D1, D2, ..., D7はタイトルT1, T2, ..., T7と、それに対する本文A1, A2, ..., A7を持っているものとする。

#### 【0034】

クラスタリング対象情報抽出部31は、このような検索結果に対し、それぞれの文書D1, D2, ..., D7を解析し、それぞれの文書のタイトルを検出する。このクラスタリング対象情報抽出部31が行うタイトルの検出は、具体的には次のようにして行う。

#### 【0035】

まず、第1の方法として、文書構造様式によってタイトルと規定される部分が

あればその部分をタイトルとする。また、第2の方法として、文書構造様式によって、標準より大きな文字で表示する指定がなされている部分があれば、その部分をタイトルとする。また、第3の方法として、定められた数の文または単語を文書先頭より抽出し、その抽出した部分をタイトルとする。さらには、これら第1、第2、第3の方法を順次行い、第1の方法を行ったとき、タイトルと規定されている部分があればその部分をタイトルとし、タイトルと規定される部分が存在しなければ、第2の方法を行い、標準より大きな文字で表示する指定がなされている部分があれば、その部分をタイトルとし、標準より大きな文字で表示する指定がなされていなければ、第3の方法を行ってタイトルを抽出する。

#### 【0036】

形態素解析部32は、クラスタリング対象情報抽出部31でそれぞれの文書からタイトルとして抽出された部分を形態素解析する。

#### 【0037】

クラスタリング処理部33は、図3に示すように、特徴要素抽出部331、特徴テーブル作成部332、文書分類部333、分類結果記憶部334、出力制御部335、表示部336などを有している。特徴要素抽出部331は、形態素解析部32で形態素解析された結果から特徴要素を抽出する。

#### 【0038】

特徴テーブル作成部332は、特徴要素抽出部331で抽出された特徴要素とそれぞれの文書D1～D7との関係を示す特徴テーブルを作成する。なお、この特徴テーブルの具体的な内容については後述する。

#### 【0039】

文書分類部333は、上述の特徴テーブルの内容を参照し、文書D1, D2, . . . , D7を意味的に共通性のある複数のクラスタに分類する。つまり、文書D1, D2, . . . , D7のそれぞれのタイトルT1, T2, . . . , T7に存在する特徴要素に基づいて、共通する特徴要素を持つ文書を1つのまとまりとし、そのまとまりを1つのクラスタとする。なお、この文書分類部333は同義特徴辞書（図示せず）を有し、共通する特徴要素を持つ文書を1つのまとまりとする処理を行う際、共通する特徴要素であるか否かの判断を、その同義語辞書を用



いて同義語が有るか否かにより行い、同義語が存在する場合にはそれを同じクラスタとするというようなことを行うことも可能である。

【0040】

分類結果記憶部 334 は、文書分類部 333 によって分類された内容を記憶する。出力制御部 335 は分類結果記憶部 334 の内容を読み出して分類結果として表示部 336 に表示させる。

【0041】

このような構成において、本発明の情報分類処理について説明する。本発明が行う情報分類処理手順は概略的には、図 4 のフローチャートに示すように、まず、汎用の検索エンジンで検索された検索結果を取得し（ステップ S1）、取得した検索結果に対しクラスタリング処理を施し（ステップ S2）、そのクラスタリング処理結果を出力する（ステップ S3）。以下、具体例を参照しながら詳細に説明する。

【0042】

ここでは、ユーザの入力したキーワードによって図 2 で示した文書 D1, D2, ..., D7 が検索結果として検索サービス 1 から出力されたとする。この検索結果はファイル形式で出力され、変換モジュール 2 でクラスタリングモジュール 3 が処理できる形式に変換されたのち、クラスタリングモジュール 3 に与えられる。

【0043】

クラスタリングモジュール 3 に入力されたこれらの文書 D1, D2, ..., D7 は、クラスタリング対象情報抽出部 31 にてタイトルが抽出される。たとえば、文書 D1 についてはタイトル T1 が検出され、文書 D2 についてはタイトル T2 が検出され、文書 D3 についてはタイトル T3 が検出されるというように、それぞれの文書 D1, D2, ..., D7 のタイトル T1, T2, ..., T7 が抽出される。

【0044】

そして、形態素解析部 32 によってそれぞれのタイトル T1, T2, ..., T7 が形態素解析されたのち、その形態素解析結果がクラスタリング処理部 33

に与えられる。クラスタリング処理部 33 では、特徴要素抽出部 331 によって、形態素解析部 32 からの形態素解析結果に基づいて、それぞれのタイトル T1, T2, ..., T7 に存在する特徴要素が抽出される。

#### 【0045】

そして、特徴テーブル作成部 332 により、それぞれの特徴要素とその特徴要素をタイトルに含む文書との関係を示す特徴テーブルが作成される。この特徴テーブルの例を図 5 に示す。なお、ここでは、文書数が 3 つ以上取り出される特徴要素とその特徴要素をタイトルに含む文書との関係を示し、特徴テーブル内に示される数値は、その特徴要素が各文書のタイトルに幾つ含まれるているかの数を示している。たとえば、「用紙」という特徴要素は、文書 D1, D4, D6, D7 のタイトル T1, T4, T6, T7 に、それぞれ 1 個ずつ含まれていることを示している。

#### 【0046】

図 5 の特徴テーブルからもわかるように、「用紙」という特徴要素をタイトルに含む文書は、文書 D1, D4, D6, D7 であり、また、「カセット」という特徴要素をタイトルに含む文書は、文書 D1, D4, D7 であり、さらに、「増設」という特徴要素をタイトルに含む文書は、文書 D2, D3, D5, D7 である。なお、先に説明した図 2 において、これら各特徴要素部分にはアンダーラインが施されている。

#### 【0047】

そして、文書分類部 333 はこのような特徴テーブルを参照して、それぞれの特徴要素ごとのクラスタ分けを行う。その分類結果を図 6 に示す。

#### 【0048】

このような分類結果は分類結果記憶部 334 に格納される。図 6 に示される分類結果において、たとえば、「用紙」で分類されたクラスタ（文書 D1, D4, D6, D7 が含まれる）について見れば、図 2 の文書内容からもわかるように、文書 D1 は用紙カセットについての内容であり、文書 D4 は用紙設定についての内容であり、文書 D6 は印刷された後の用紙の汚れについての内容であり、文書 D7 は用紙カセットの増設についての内容である。

## 【0049】

このように、これらの文書D1, D4, D6, D7はどれも用紙に関する内容であり、1つのクラスタとして分類されて何等问题のないものとなり、その分類結果は適切であるといえる。

## 【0050】

また、「カセット」で分類されたクラスタ（文書D1, D4, D7が含まれる）について見れば、図2の文書内容からもわかるように、文書D1は用紙カセットについての内容であり、文書D4は用紙設定についての内容であり、文書D7は用紙カセットの増設についての内容である。

## 【0051】

また、このように、これらの文書D1, D4, D6, D7にはどれも用紙をセットすることに関する内容が含まれており、1つのクラスタとして分類されて何等问题のないものとなり、その分類結果は適切であるといえる。

## 【0052】

また、「増設」で分類されたクラスタ（文書D2, D3, D5, D7が含まれる）について見れば、図2の文書内容からもわかるように、文書D2はメモリの増設についての内容であり、文書D3はインタフェースカードの増設についての内容であり、文書D5はハードディスクの増設についての内容であり、文書D7は用紙カセットの増設についての内容である。

## 【0053】

このように、これらの文書D2, D3, D5, D7はどれも何かを増設する場合についての内容であり、1つのクラスタとして分類されて何等问题のないものとなり、その分類結果は適切であるといえる。

## 【0054】

このような適切な分類が行える理由としては、それぞれの文書のタイトルから特徴要素を抽出し、その特徴要素に基づいて文書を分類しているからである。つまり、文書のタイトルは、その文書の作成者がその文書の主旨を表す内容を表現していることが多い。したがって、文書のタイトルに含まれる特徴要素を用いて分類を行うことにより、分類結果が散漫になることが少なく、また、ノイズクラ

スタが生成される率も少なくすることができる。また、各文書のタイトルは、その文書の作成者がその文書の主旨を表す内容を表現していることから、文書の制作者側の視点による分類が得られる。

【0055】

なお、ユーザに実際に提示されるクラスタリング結果は、たとえば図7に示すような内容であり、それぞれの特徴要素とその特徴要素を有するタイトルの一覧が提示される。そして、ユーザはこのようなクラスタリング結果の一覧を見て、自分の欲しい情報の入っているようなタイトル部分をクリックすれば、そのタイトルに対応する本文が表示されるというような表示処理がなされる。

【0056】

このように、この実施の形態では、ユーザがある汎用の検索サービスを利用し、その検索サービスに何らかのキーワードを入力することによって、複数の文書D1, D2, ..., D7が検索された場合、これら複数の文書D1, D2, ..., D7のタイトルT1, T2, ..., T7を抽出し、そのタイトルに含まれる特徴要素に基づいて、それぞれの文書D1, D2, ..., D7に対しクラスタリング処理を行う。

【0057】

これによって、従来では、検索サービスによって検索された検索結果は単に羅列的に提示されるだけであったものが、本発明によれば、検索された検索結果のタイトル内容に基づいたクラスタリング結果をユーザに提示できる。そのクラスタリング結果（図7参照）は、それぞれのタイトルに含まれる特徴要素ごとにそれぞれのタイトルが分類された内容となり、ユーザが見易いように整理されたものとなる。

【0058】

そして、その中から見たい情報があれば、タイトル部分をクリックすれば、それぞれのタイトルを有する文書を表示させることができる。

【0059】

また、これまでの説明では、ある1つの汎用の検索サービスで検索された結果をクラスタリング処理する場合について説明したが、複数の検索サービスにより

検索された結果をクラスタリング処理する場合にも適用できる。

【0060】

検索サービスはそれぞれに得意の分野があることも多く、たとえば、ある検索サービスはスポーツ関係の情報を多数保有し、ある検索サービスは学術関係の情報を多数保有し、また、ある検索サービスは芸能関係の情報を多数保有しているというように、それぞれの得意の分野が存在する場合も多い。これらそれぞれの得意分野については豊富な情報を所有しており、ユーザの所望とする情報が適切に取り出される可能性が高い。したがって、情報検索を行う際は、検索サービスを使い分けることも普通に行われる。このように、複数の検索サービスを用いる場合のクラスタリング処理について以下に説明する。

【0061】

図8は複数の検索サービスを用いてこれまで説明したクラスタリング処理を行うための情報分類装置の構成を説明する図であり、ここでは、複数の検索サービスとして第1の検索サービス1a、第2の検索サービス1b、第3の検索サービス1cの3つの検索サービスが存在するものとする。

【0062】

このように複数の検索サービス（ここでは検索サービス1a、1b、1c）を対象とする場合には、それぞれの検索サービスにより検索された検索結果の内容、長さ、検索結果出力順序などがまちまちなので、それぞれの検索サービス1a、1b、1cからのファイルをクラスタリングモジュール3で処理可能な形式に変換する変換モジュール2a、2b、2cを検索サービス1a、1b、1cに対応して用意する。なお、クラスタモジュール3の構成は図1と同じであるので同一部分には同一符号が付されている。

【0063】

このような構成であれば、ユーザは検索しようとする情報の分野に応じて検索サービスを使い分けることができる。たとえば、第1の検索サービス1aがスポーツ関係の情報検索に向いている検索サービスであるとすれば、スポーツ関係の情報を検索しようとする場合、第1の検索サービス1aを用いて検索を行う。また、第2の検索サービス1bが学術関係の情報検索に向いている検索サービスで

あるとすれば、学術関係の情報を検索しようとする場合、第2の検索サービス1bを用いて検索を行う。

【0064】

このように、ユーザは検索しようとする情報によって検索サービスを選択することができ、しかも、それによって検索された結果は、クラスタリングモジュール3によってクラスタリング処理がなされ、ユーザが見やすいように整理された状態で提示できる。なお、このクラスタリング処理については、すでに説明したのでここではその説明は省略する。

【0065】

このように複数の検索サービスを選択的に使用できるようにすれば、それぞれの得意分野に応じた検索が可能となるばかりでなく、ある1つの検索サービスが混み合っているような場合には、他の検索サービスに切り換えて検索を行うというような柔軟な検索も可能となる。

【0066】

さらに、複数の検索サービスによる検索処理を並行に行い、それぞれの検索サービスによる検索結果を1つにまとめて、まとめられた内容についてクラスタリング処理することも可能である。これを図9により簡単に説明する。

【0067】

図9の構成は図8の構成において、第1～第3の検索サービス1a, 1b, 1cに対応して設けられた変換モジュール2a, 2b, 2cとクラスタリングモジュール3との間に、それぞれの検索結果を収集して1つにまとめる検索結果収集部4を設けた構成となっている。その他は、図8と同じ構成であり、同一部分には同一符号が付されている。

【0068】

このような構成であれば、複数の検索サービス（ここでは第1～第3の検索サービス1a, 1b, 1c）では、ユーザの入力したキーワードに対して並行して検索処理を行い、それぞれの検索サービス1a, 1b, 1cによって検索されたそれぞれの検索結果を、それぞれの検索サービス1a, 1b, 1cに対応する変換モジュール部2a, 2b, 2cでクラスタリングモジュール3が処理可

能な形式に変換し、それぞれの変換後のファイルを検索結果収集部4に与え、それぞれの検索結果を1つにまとめる。そして、そのまとめられた検索結果は、クラスタリングモジュール3に入力され、これまで説明したようなクラスタリング処理を行う。

## 【0069】

このように、複数の検索サービスを用いて検索処理を行うことにより、1つの検索サービスでは検索できなかった幅広い情報を取得することも可能である。また、検索範囲が広がるので、網羅的に情報を探すことができ、どのような情報が世の中に存在するのかを見極めるのに非常に有効となる。このようにして得られた情報は、これまで説明したようなクラスタリング処理がなされ、見やすい形でユーザに提示される。

## 【0070】

また、この図9に示した第1～第3の検索サービス1a, 1b, 1cを用いて検索処理を並行して行うような場合、それぞれの検索サービス1a, 1b, 1cで得られた検索結果（変換モジュール2a, 2b, 2c出力）を1つにまとめたあと、クラスタリング処理を行うのではなく、それぞれの検索サービス1a, 1b, 1cで得られた検索結果（変換モジュール2a, 2b, 2c出力）それぞれについてクラスタリング処理を行い、それぞれのクラスタ処理結果をユーザに提示するようにしてもよい。

## 【0071】

これによって、広範囲にわたって存在する情報を効率よく検索することが可能となり、また、ユーザは第1～第3の検索サービス1a, 1b, 1cによるそれぞれの検索結果のそれぞれのクラスタリング結果を見比べたりすることができ、それぞれの検索サービスの特色などを知ることができる。

## 【0072】

なお、本発明は以上説明した実施の形態に限定されるものではなく、本発明の要旨を逸脱しない範囲で種々変形実施可能となるものである。たとえば、上述の実施の形態では、クラスタリングを行うための情報（クラスタリング対象情報）として、検索されたそれぞれの文書のタイトルを用いた例について説明したが、

これは、タイトルだけでなく、たとえば、URLアドレス（http://を取り除いた部分）、更新日時（単純な時間または最近1カ月以内の更新日時）、ファイルサイズ（webページ本文のバイトサイズなど）を用いてクラスタリングすることもできる。また、これらは、単独で用いてクラスタリングするようにしてもよく、幾つかを組み合わせてもよい。このように、クラスタリング対象情報を種々選ぶことによって、それぞれに応じた特色のあるクラスタリングが行える。そして、これらのどれを用いるかは、最初にメニューなどで選択項目を選ぶことで可能となる。また、選んだ項目が無い場合には、他の項目を代用する。たとえば、タイトルを選んだ場合、webページにタイトルが無い場合には、URLアドレスを代用する。

#### 【0073】

また、以上説明した本発明の情報分類処理を行う処理プログラムは、フロッピーディスク、光ディスク、ハードディスクなどの記録媒体に記録させておくことができ、本発明はその記録媒体をも含むものである。また、ネットワークから処理プログラムを得るようにしてもよい。

#### 【0074】

##### 【発明の効果】

以上説明したように本発明によれば、汎用の検索サービスで検索された多数の検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を提示するようにしているので、従来の検索結果がそのまま羅列的に出力されるものと異なり、ユーザが見易いように整理されたものとなる。なお、この場合、検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュール部を備える。そして、その変換モジュールを複数の検索サービスそれぞれに対応して用意することによって、複数の検索サービスで検索された検索結果に対してクラスタリング処理が可能となる。

#### 【0075】

たとえば、複数の検索サービスのいずれかを選択し、選択された検索サービスで検索された検索結果に対してクラスタリング処理が可能となる。これによって、ユーザは、情報を検索使用とする際、検索しようとする情報の種類によって検



索サービスを選ぶことができる。さらに、1つの検索サービスが混雑している場合、他の検索サービスに切り換えて検索を行うということも可能となる。

【0076】

また、複数の検索サービスの少なくとも2つの検索サービスを用いて並行して検索処理を行い、複数の検索サービスで得られたそれぞれの検索結果を収集して、その収集された検索結果に対してクラスタリング処理を行うことができる。これによって、広範囲にわたって存在する情報を効率よく検索することが可能となり、その広範囲にわたる情報から検索された検索結果を1つにまとめ、そのまとめられた結果をクラスタリングすることができ、多種多様な情報のクラスタリング結果を提示できる。

【0077】

さらに、複数の検索サービスの少なくとも2つの検索サービスを用いて検索処理を行い、それぞれの検索サービスで検索された情報を個々にクラスタリングしてそれぞれのクラスタリングを提示することも可能で、これによれば、広範囲にわたって存在する情報を効率よく検索することが可能となり、また、それぞれのクラスタリング結果を比較することが可能となり、それぞれの検索サービスの特色などを知ることができる。

【0078】

また、クラスタリング対象とする情報は、それぞれの検索結果における文書のタイトル、URLアドレス、更新日時、ファイルサイズの少なくとも1つとして、いる。これによって、クラスタリングを文書全体から行う場合に比べて、クラスタリング処理を簡単なものとすることができる。それぞれに特色を有したクラスタリング結果が得られる。

【図面の簡単な説明】

【図1】

本発明の情報分類装置の実施の形態を説明する図であり、ある1つの検索サービスで得られた検索結果に対してクラスタリングを行う場合の情報分類装置の構成を説明するブロック図である。

【図 2】

本発明の実施の形態で用いられるある検索サービスで検索された検索結果としての複数の文書例を示す図である。

【図 3】

図 1 で示されたクラスタリング処理部の構成を示すブロック図である。

【図 4】

本発明の文書分類処の手順を概略的に説明するフローチャートである。

【図 5】

図 2 で示されたそれぞれの文書のタイトルから抽出された特徴要素とその特徴要素をタイトルに含む文書との関係を示す特徴テーブルの内容の示す図である。

【図 6】

図 5 の特徴テーブルを基にしてそれぞれの文書を分類した分類結果を示す図である。

【図 7】

図 6 の分類結果に基づいてそれぞれの文書のタイトルをクラスタリングした例を示す図である。

【図 8】

複数の検索サービスから得られた検索結果に対してクラスタリングを可能とする場合の情報分類装置の構成を示すもので、選択されたある 1 つの検索サービスから得られた検索結果に対してクラスタリングを行う場合の情報分類装置の構成を説明するブロック図である。

【図 9】

複数の検索サービスから得られた検索結果に対してクラスタリングを可能とする場合の情報分類装置の構成を示すもので、複数の検索サービスで並行して検索を行い、それぞれから得られた検索結果に対してクラスタリングを行う場合の情報分類装置の構成を説明するブロック図である。

【符号の説明】

- 1 検索サービス
- 2 変換モジュール

3 クラスタリングモジュール

4 検索結果収集部

3 1 クラスタリング対象情報抽出部

3 2 形態素解析部

3 3 クラスタリング処理部

3 3 1 特徴要素抽出部

3 3 2 特徴テーブル作成部

3 3 3 文書分類部

3 3 4 分類結果記憶部

3 3 5 出力制御部

3 3 6 表示部

1 a, 1 b, 1 c 第1～第3の検索サービス

2 a, 2 b, 2 c 第1～第3の変換モジュール

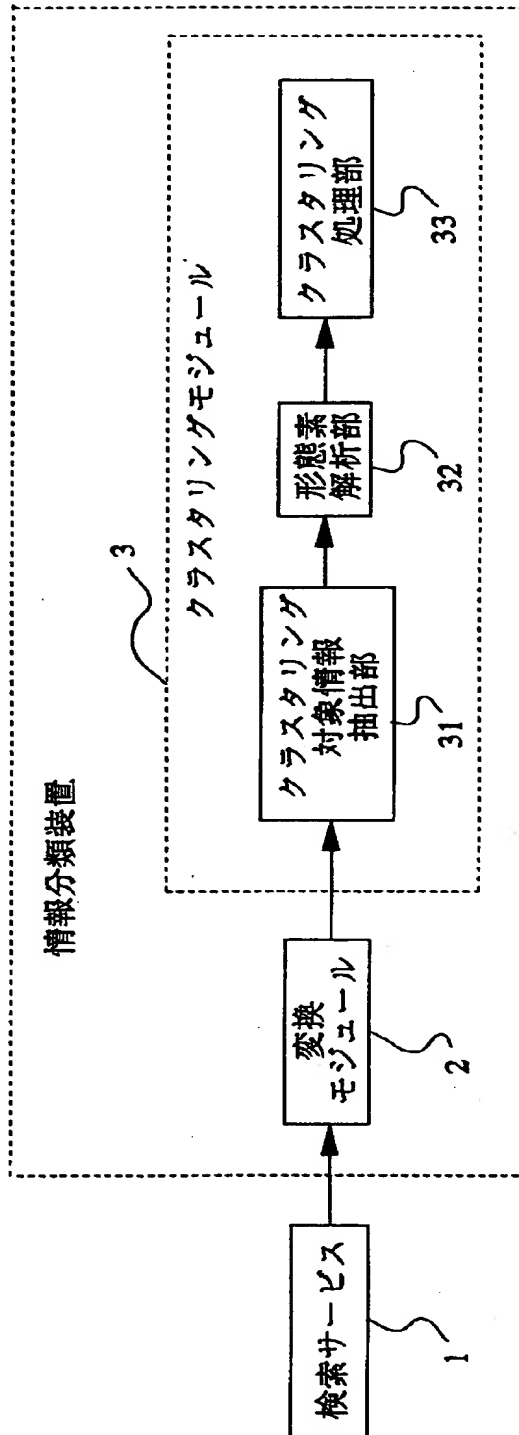
A 1, A 2, . . . , A 7 本文

D 1, D 2, . . . , D 7 文書

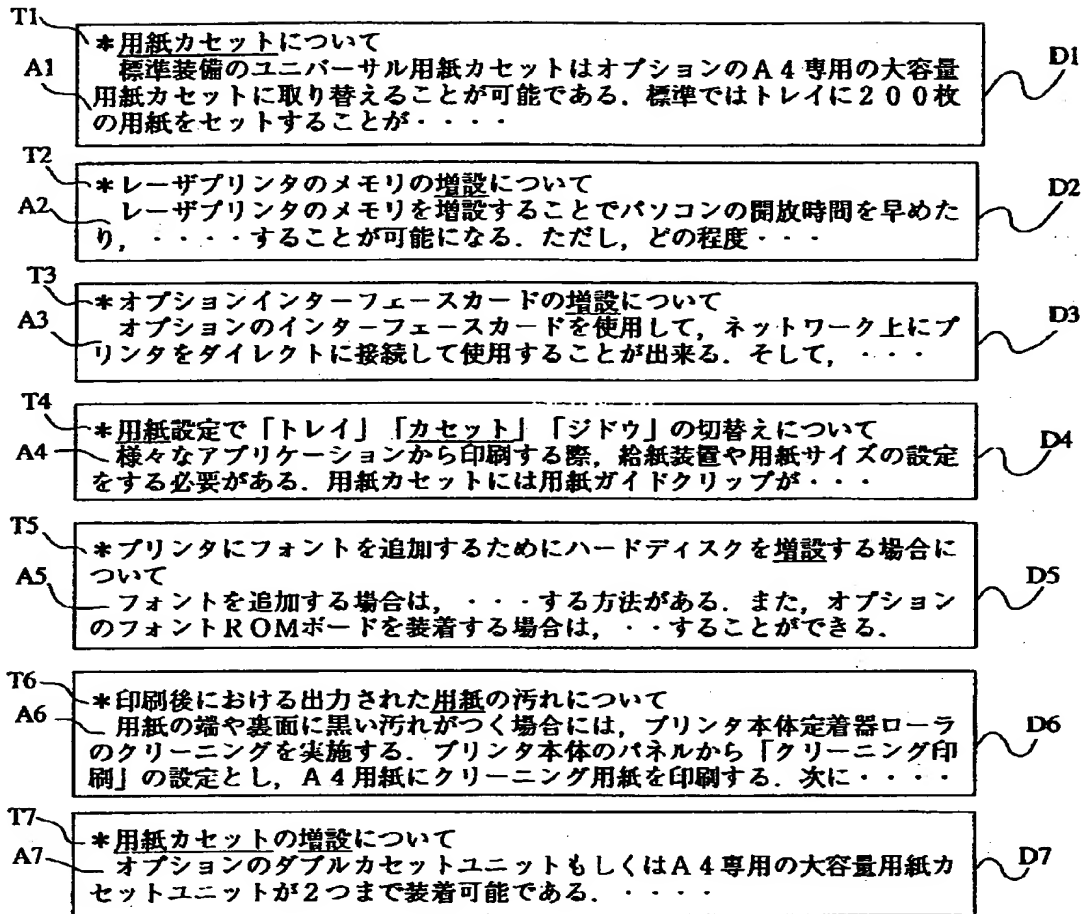
T 1, T 2, . . . , T 7 タイトル

【書類名】 図面

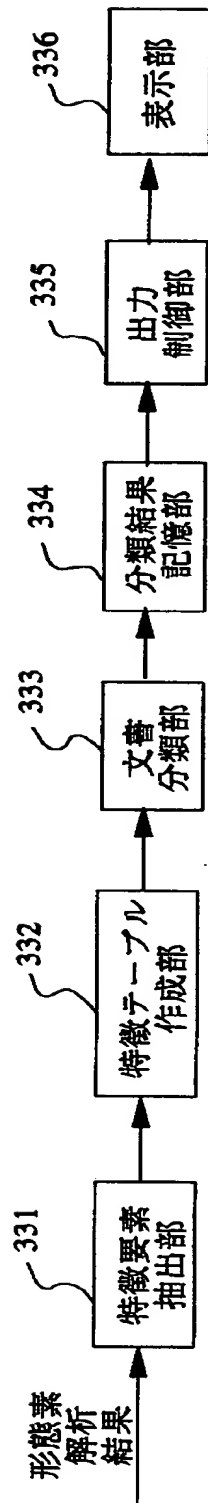
【図 1】



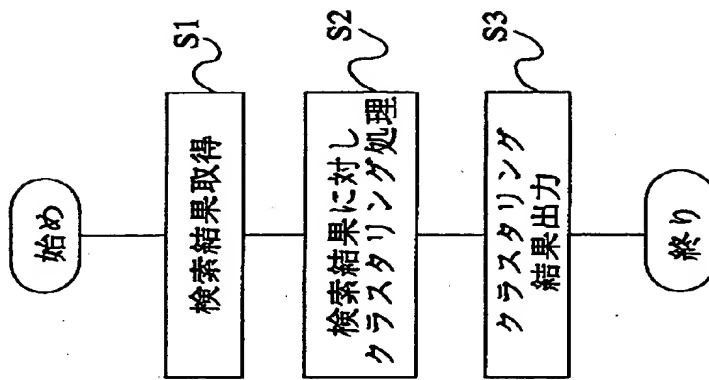
【図 2】



【図 3】



【図 4】



【図 5】

特徴要素	文書 D 1	文書 D 2	文書 D 3	文書 D 4	文書 D 5	文書 D 6	文書 D 7
用紙	1			1		1	1
カセット	1			1			1
増設		1	1		1		1



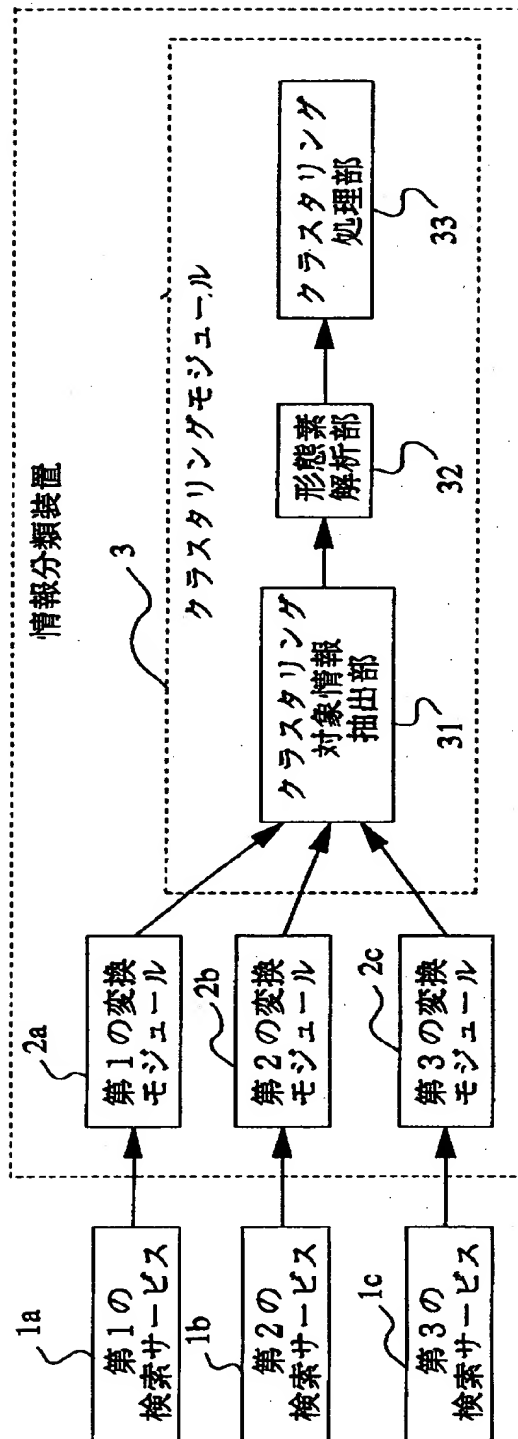
【図6】

特徴要素	クラスタ
用紙	D1, D4, D6, D7
カセット	D1, D4, D7
増設	D2, D3, D5, D7

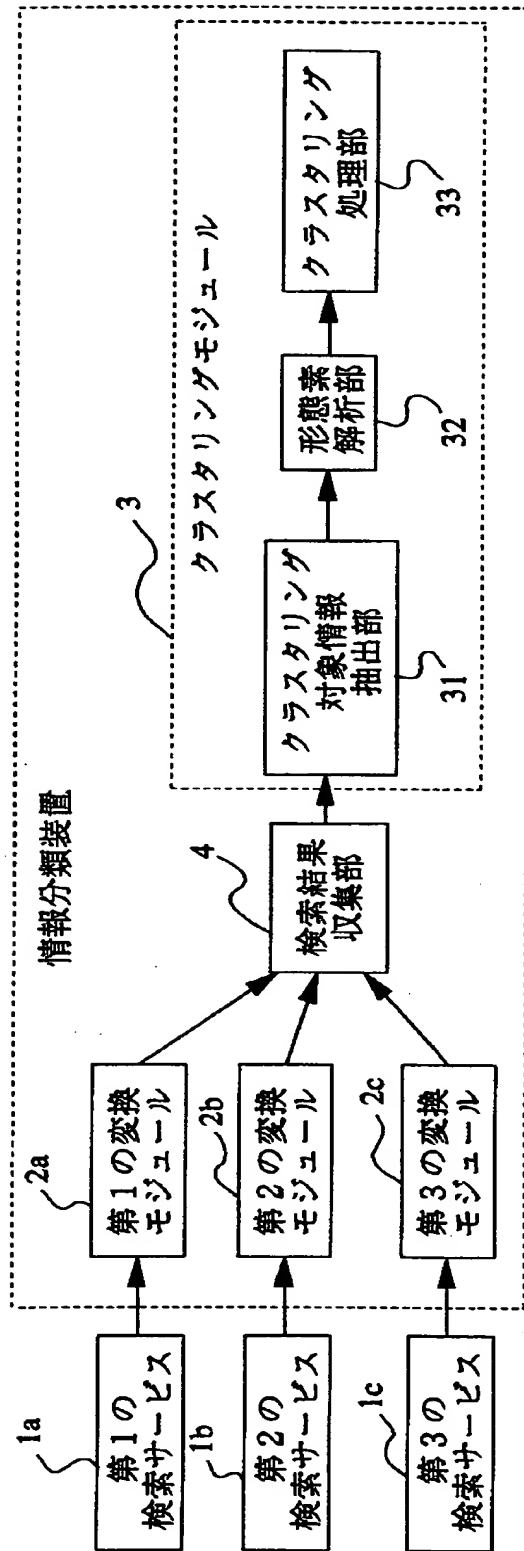
【図 7】

特徴要素	タイトル
用紙	<ul style="list-style-type: none"> <li>* 用紙カセットについて</li> <li>* 用紙設定で「トレイ」「カセット」「ジドウ」の切替えについて</li> <li>* 印刷後における出力された用紙の汚れについて</li> <li>* 用紙カセットの増設について</li> </ul>
カセット	<ul style="list-style-type: none"> <li>* 用紙カセットについて</li> <li>* 用紙設定で「トレイ」「カセット」「ジドウ」の切替えについて</li> <li>* 用紙カセットの増設について</li> </ul>
増設	<ul style="list-style-type: none"> <li>* レーザープリンタのメモリ増設について</li> <li>* オプショントウェアカードの増設について</li> <li>* プリンタにフォントを追加するためにハードディスクを増設する場合について</li> <li>* 用紙カセットの増設について</li> </ul>

【図 8】



【図9】



【書類名】 要約書

【要約】

【課題】 汎用の検索サービスを用いて検索を行う場合、ユーザの入力するキーワードに対して検索された多数の情報は羅列的に提示されるだけであり、検索結果が多いと検索結果の内容の把握がしにくい。

【解決手段】 汎用の検索サービス 1 で検索された複数の検索結果をクラスタリングモジュール 3 が取得して、そのクラスタリングモジュール 3 では、その検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を提示する。このとき、検索サービス 1 で検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュール 2 を備える。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000002369]

1. 変更年月日	1990年 8月20日
[変更理由]	新規登録
住 所	東京都新宿区西新宿2丁目4番1号
氏 名	セイコーエプソン株式会社